GENES INVOLVED IN POLYKETIDE SYNTHASE PATHWAYS AND USES THEREOF

BACKGROUND OF THE INVENTION

Technical Field

The subject invention relates to isolated nucleic acid sequences or genes involved in polyketide synthase (PKS) biosynthetic pathways. In particular, such pathways are involved in the production of polyunsaturated fatty acids (PUFAs) such as, for example, Eicosapentaenoic acid (EPA) and Docosahexaenoic acid (DHA). Specifically, the invention relates to isolating nucleic acid sequences encoding proteins involved in eukaryotic PUFA-PKS systems and to uses of these genes and encoded proteins in PUFA-PKS systems, in heterologous hosts, for the production of PUFAs such as EPA and DHA.

Background Information

Long chain polyunsaturated fatty acids (PUFAs) that contain 20 or 22 carbon atoms (C_{20}-, C_{22}-PUFAs) are essential components of membrane phospholipids and serve as precursors of eicosanoids like prostaglandin, leukotrienes and thromboxanes. They also play a pivotal role in various biological functions such as fetal growth and development, retina functioning and the inflammatory response. The n-6 fatty acids and the n-3 fatty acids are the two major classes of long chain PUFAs. In mammals, the major endpoint of the n-6 pathway is arachidonic acid (ARA, 20:4n-6), and the major endpoints of the n-3 pathway are eicosapentaenoic acid (EPA, 20:5n-3) and docosahexaenoic acid (DHA, 22:6n-3). n-6 and n-3 PUFAs are metabolically and functionally distinct, quite often having opposing physiological functions; thus, their balance is important for homeostasis. An excess of n-6 PUFAs shifts the physiological state to one that is

prothrombotic and preaggregratory, leading to inflammatory and cardiovascular complications. On the other hand, n-3 PUFAs such as EPA and DHA have been shown to have therapeutic value in prevention and treatment of diseases such as, for example, cardiovascular disease, inflammation, arthritis and cancer. Thus, there is interest in identifying inexpensive and renewable sources of EPA and DHA.

A large number of lower eukaryotes like fungi and algae produce long chain PUFAs such as EPA and DHA. The exact mechanism of PUFA biosynthesis in these organisms is unknown but is presumed to be similar to that of mammals (i.e., an aerobic pathway involving an alternating series of desaturations and elongations catalyzed by a series of enzymes called desaturases and elongases). Many of these enzymes have already been identified in several of these PUFA-rich fungi such as *Thraustochytrium* sp., *Mortierella* sp., etc (Knutzon et al., *J. Biol. Chem* (1998) 273:29360-29366; Parker-Barnes et al., *Proc. Natl. Acad. Sci. USA.* (2000) 97:8284-8289; Huang et al., *Lipids* (1999) 34:649-659; Qiu et al., *J. Biol. Chem.* (2001) 276:31561-31566).

Recently, Metz et al. (*Science* (2001) 293: 290-293) proposed that DHA biosynthesis in *Schizochytrium*, an organism that belongs to the *Thraustochytrid* family, occurs via a novel polyketide synthase (PKS) pathway rather than the desaturase/elongase pathway (see also U.S. Patent No. 6,566,583). This mechanism is thought to be similar to that used for EPA/DHA production in prokaryotes like *Shewanella* (Yazawa, *Lipids* (1996) 31 Suppl: S297-300) and *Vibrio* (Morita et al., *Biotechnol. Lett.* (1999) 21:641-646). In particular, PUFA production is initiated by the condensation between a short chain starter unit like acetyl CoA and an extender unit like

malonyl CoA. The C4 acyl chain formed is covalently attached to an acyl carrier protein (ACP) domain of the PKS complex and goes through successive rounds of reduction, dehydration, reduction, and condensation, with the acyl chain growing by C2 units with each round. A novel dehydratase/isomerase has been proposed to exist in the complex (Metz et al., *Science* (2001) 293:290-293) that can catalyze trans- to cis- conversion of the double bonds, thus generating double bonds in the correct position of EPA and DHA.

The genes involved in the PUFA-PKS pathway have been identified from a number of marine organisms including *Shewanella*. In *Shewanella*, these genes were arranged in five open reading frames (ORFs) of ~ 20 kb in length and were shown to be sufficient for EPA production when tested in *E. coli* (Yazawa, *Lipids* (1996) 31 Suppl: S297-300). Examination of the protein sequences encoded by these five ORFs revealed that at least eleven enzymatic domains could be identified, seven of which were more strongly related to PKS proteins (Metz et al., *Science* (2001) 293:290-293) rather than to the fatty acid synthase (FAS) proteins that were suggested earlier (Watanabe et al., *J. Biochem.* (1997) 122:467).

It has been suggested that in *Shewanella*, at least some of the double bonds are introduced into EPA by a dehydratase-isomerase mechanism catalyzed by the fabA-like domain present in ORF 7 of the *Shewanella* PUFA-PKS cluster (Metz et al., *Science* (2001) 293:290-293). Expression studies of the *Shewanella* PKS gene cluster in *E.coli* revealed that EPA production could take place in the absence of oxygen indicating that the aerobic desaturase pathway did not play any role in EPA production in these marine bacteria. Thus, PUFA production in this marine bacteria is thought to occur

via a novel PKS-like pathway and this is thought to be widespread in marine bacteria that make PUFAs, since genes with high homology to the *Shewanella* PUFA-PKS gene cluster have been identified in *Vibrio marinus* (Tanaka et al., *Biotechnol. Lett.* (1999) 21:939) and in *Photobacterium profundum* (Allen et al., *Appli. Environ. Microbiol.* (1999) 65:1710). The PKS pathways for PUFA synthesis in *Shewanella* and *Vibrio marinus* have been described in U.S. Patent No. 6,140,486.

Genes homologous to the *Shewanella* PUFA-PKS gene cluster were recently identified in *Schizochytrium*, a marine eukaryote that produces DHA (Metz et al., *Science* (2001) 293: 290-293; see also U.S. Patent No. 6,566,583). Labeling experiments with *Schizochytrium* demonstrated that DHA was produced solely from an acetate precursor, rather than from any C_{18} fatty acid intermediate, pointing to the PKS-PUFA pathway as being functional in DHA production rather than the aerobic desaturase pathway.

Because of the increased demand for PUFAs such as EPA and DHA, alternate sources of these PUFAs are being sought after. The current natural sources of n-3 PUFAs such as fish oil are not economical or renewable and thus not suitable for commercial needs. Thus, the development of transgenic plant oils enriched with ω-3 PUFAs is currently being considered. For this, the plant will need to be genetically engineered to contain desaturase and elongase genes that are involved in EPA/DHA production. However, this would require expression of six to seven separate enzymes simultaneously in plants, and further manipulations might be necessary to control the flux through the pathway, target these genes to specific organelles, and/or modulate gene expression so as to prevent the accumulation of undesirable intermediates. Thus, it would be of interest to identify

alternate PUFA biosynthesis pathways such as the PUFA-PKS pathway.

Although the bacterial PUFA-PKS genes do provide a novel resource for producing transgenic plant oils, it is not known how these bacterial genes will function in a eukaryotic host. Also, the source organisms for these genes grow in cold marine environments and their enzyme systems might not function well at or above 30°C which could pose a problem for expression in some crops. Additionally, the PUFAs in these marine bacteria are not stored in the triglyceride form since these organisms are not oleaginous strains; thus, the PUFA-PKS system in these organisms cannot direct triglyceride formation. These shortcomings may be overcome by identifying additional PUFA-PKS genes from eukaryotic sources that make triglycerides. The identification of a PUFA-PKS gene cluster from *Schizochytrium*, fits this criteria. However, the amount of DHA produced by *Schizochytrium* is low compared to other *Thraustochytrid* species, and a large fraction of this DHA is found in the phospholipid fraction rather than in the triglyceride form (Kendrick et al., *Lipids* (1992) 27:15-20). Therefore, there is a need to identify other PUFA-PKS systems from eukaryotes that produce large amounts of DHA that is found in the triglyceride fraction, as well as EPA. *Thraustochytrium aureum* is an ideal candidate since this organism belongs to the same *Thraustochytrid* family as *Schizochytrium* does, but produces copious amounts of DHA (~ 30% of the total lipid is DHA) as compared to *Schizochytrium*, and has a major portion of its DHA in the triglycerol fraction (Kendrick et al., *Lipids* (1992) 27:15-20). Identification of the PUFA-PKS system from *Thraustochytrium aureum* provides an excellent alternative for the production of PUFA-enriched transgenic oils.

All U.S. patents and publications referred to herein are hereby incorporated in their entirety by reference.

SUMMARY OF THE INVENTION

The present invention encompasses an isolated nucleic acid sequence or fragment thereof comprising or complementary to a nucleic acid sequence encoding a polypeptide, wherein the amino acid sequence of said polypeptide has at least 65% amino acid identity to an amino acid sequence comprising SEQ ID NO:10.

Additionally, the present invention includes an isolated nucleic acid sequence or fragment thereof comprising or complementary to a nucleic acid sequence having at least 70% nucleotide sequence identity to a nucleic acid sequence comprising SEQ ID NO:8.

Further, the invention also encompasses an isolated nucleic acid sequence or fragment thereof comprising or complementary to a nucleic acid sequence encoding a polypeptide, wherein the amino acid sequence of said polypeptide has at least 65% identity to an amino acid sequence comprising SEQ ID NO:11.

Also, the present invention includes an isolated nucleic sequence or fragment thereof comprising or complementary to a nucleic acid sequence having at least 70% nucleotide sequence identity to a nucleic acid sequence comprising SEQ ID NO:9. Each of the nucleic acid sequences referred to above encodes a functionally active polyketide synthase enzyme. This enzyme modulates the production of at least one polyunsaturated fatty acid (PUFA) when expressed in a host cell. The PUFA may be, for example, eicosapentaenoic acid or docosahexaenoic acid. Further, each of the nucleic acid sequences may be isolated from, for example, *Thraustochytrium sp.* and, in particular, from *Thraustocytrium aureum*. The present

invention also includes a protein or polypeptide encoded by any one or more of the above-described nucleic acid sequences or fragments thereof.

Additionally, the present invention also encompasses a purified protein or fragment thereof comprising an amino acid sequence having at least 65% amino acid identity to an amino acid sequence comprising SEQ ID NO:10 or SEQ ID NO:11.

Further, the invention includes a method of producing a polyketide synthase enzyme. This method comprises the steps of isolating a nucleic acid sequence comprising SEQ ID NO:8 or SEQ ID NO:9; constructing a vector comprising the isolated nucleic acid sequence operably linked to a regulatory sequence; and introducing the vector into a host cell under time and conditions sufficient for expression of the polyketide synthase enzyme. The host cell may be either a eukaryotic cell or a prokaryotic cell.

The present invention also encompasses a vector comprising a nucleic sequence comprising SEQ ID NO:8 or SEQ ID NO:9, operably linked to a regulatory sequence as well as a host cell comprising this vector. Again, the host cell may be either a eukaryotic cell or a prokaryotic cell.

Moreover, the present invention also includes a plant cell, plant or plant tissue comprising the above-described vector, wherein expression of the nucleic acid sequence of the vector results in production of at least one polyunsaturated fatty acid by the plant cell, plant or plant tissue. The at least one polyunsaturated fatty acid may be, for example, eicosapentaenoic acid (EPA) or docosahexaenoic acid (DHA). The invention also includes one or more plant oils or acids expressed by the plant cell, plant or plant tissue described above.

Additionally, the present invention includes a transgenic plant comprising the above-described vector, wherein expression of the nucleic acid sequence of the vector results in production of at least one polyunsaturated fatty acid in seeds of the transgenic plant.

Further, the present invention also includes a method for producing a polyunsaturated fatty acid. This method comprises the steps of isolating a nucleic acid sequence comprising SEQ ID NO:8 or SEQ ID NO:9; constructing a vector comprising the isolated nucleic acid sequence operably linked to a regulatory sequence; introducing the vector into a host cell for a time and under conditions sufficient for expression of a polyketide synthase enzyme encoded by the isolated nucleic acid sequence; exposing the polyketide synthase enzyme to a substrate to produce a product; and exposing the product to at least one enzyme selected from the group consisting of a ketosynthase, a ketoreductase, a dehydratase, an isomerase, an enoyl reductase, a desaturase and an elongase in order to produce the polyunsaturated fatty acid. The substrate may be, for example,acetyl-CoA malonyl-CoA, malonyl-ACP, methylmalonyl-CoA or methylmalonyl-ACP. The polyunsaturated fatty acid may be, for example, EPA or DHA. The invention also includes a composition comprising at least one polyunsaturated fatty acid produced according to the above-described method. In the composition, the at least one polyunsaturated fatty acid may be, for example, EPA or DHA.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a comparison of the predicted amino acid sequence of the *Thraustochytrium aureum* probe 'TA-PKS-1-consensus' and the homologous region on ORF A of the *Schizochytrium* gene cluster (Accession number AAK72879).

Figure 2 illustrates a comparison of the predicted amino acid sequence of the *Thraustochytrium aureum* probe 'TA-PKS-1-consensus' and the homologous region on ORF 5 of the *Shewanella* PKS gene cluster (Accession number AAB81123).

Figure 3 represents the organization of ORFA and ORFB of the PUFA-PKS genes from *Thraustochytrium aureum* (ATCC 34304). (KS = ß-keto acyl synthase; MAT = MalonylCoA transferase; ACP = Acyl carrier protein; KR = Ketoacyl-ACP reductase; AT = Acyl transferase; CLF = Chain length factor; ER = Enoyl reductase; DH = Dehydratase)

Figure 4 illustrates all of the sequences and corresponding sequence identifier numbers referred to herein.

DETAILED DESCRIPTION OF THE INVENTION

The subject invention relates to isolated nucleic acid sequences or molecules (and the proteins encoded thereby) involved in PKS pathways and thus in the production of polyunsaturated fatty acids (PUFAs) such as DHA and EPA. Such PUFAs may be added to, for example, pharmaceutical and nutritional compositions. Furthermore, the subject invention also includes uses of the cDNAs and of the proteins encoded by the genes.

The Nucleic Acid Sequences of the Two Genes (Open Reading Frames A and B) and the Encoded Proteins

The nucleic acid sequence of the first isolated gene (ORF A) from *T. aureum* ATCC 34304 is shown in Figure 4 (SEQ ID NO:8), and the amino acid sequence of the encoded purified protein or enzyme encoded by this nucleic acid sequence is also shown in Figure 4 (SEQ ID NO:10). Additionally, the nucleic acid sequence of the second isolated gene (ORF B) from *T. aureum* ATCC 34304 is shown in Figure 4 (SEQ ID NO:9), and the amino acid sequence of the purified protein encoded by this nucleic acid sequence is also shown in Figure 4 (SEQ ID NO:11).

It should be noted that the present invention also encompasses nucleic acid sequences or molecules (and the corresponding encoded proteins) comprising nucleotide sequences which are at least about at least about 65% identical to, preferably at least about 70% identical to, more preferably at least about 80% identical to, and most preferably at least about 90% identical to the nucleotide sequence of SEQ ID NO:8. Further, the present invention also includes nucleic acid sequences or molecules (and the corresponding encoded proteins) comprising nucleotide sequences which are at least about 65% identical to, preferable at least about 70% identical to, more preferably at least about 80% identical to, and most preferably at least about 90% identical to the nucleotide sequence of SEQ ID NO:9. Complements of these sequences are also encompassed by the present invention. (All integers within the range of 65 to 100 (in terms of percent identity) are also included within the scope of the invention.)

The sequences having the above-described percent identity (or complementary sequences) may be derived from one or more sources other than *T. aureum* (e.g., other

eukaryotes (e.g., Thraustochytrium spp. (e.g., Thraustochytrium roseum)), Schizochytrium spp. (e.g., Schizochytrium aggregatum), Conidiobolus spp. (e.g., Conidiobolus nanodes), Entomorphthora spp. (e.g., Entomorphthora exitalis), Saprolegnia spp. (e.g., Saprolegnia parasitica and Saprolegnia diclina), Leptomitus spp. (e.g., Leptomitus lacteus), Entomophthora spp., Pythium spp., Porphyridium spp. (e.g., Porphyridium cruentum), Conidiobolus spp., Phytophathora spp., Penicillium spp., Coidosporium spp., Mucor spp. (e.g., Mucor circinelloides and Mucor javanicus), Fusarium spp., Aspergillus spp., Rhodotorula spp., Amphidinium carteri, Chaetoceros calcitrans, Cricosphaera carterae, Crypthecodinium cohnii, Cryptomonas ovata, Euglena gracilis, Gonyaulax polyedra, Gymnodinium spp. (e.g. Gymnodinium nelsoni), Gyrodinium cohnii, Isochrysis spp. (e.g. Isochrysis galbana), Microalgae MK8805, Nitzschia frustulum, Pavlova spp. (e.g., Pavlova lutheri), Phaeodactylum tricornutum, Prorocentrum cordatum, Rhodomonas lens, and Thalassiosira pseudonana), a Psychrophilic bacteria (e.g., Vibrio spp. (e.g., Vibrio marinus)) and a yeast (e.g., Dipodascopsis uninucleata.

Furthermore, the present invention also encompasses fragments and derivatives of the nucleic acid sequences of the present invention (i.e., SEQ ID NO:8 (ORF A)and SEQ ID NO:9 (ORF B)) as well as of the corresponding sequences derived from non-*T. aureum* sources, as described above, and having the above-described complementarity or identity. Functional equivalents of the above-sequences (i.e., sequences having polyketide synthase activity) are also encompassed by the present invention.

For purposes of the present invention, "complementarity" is defined as the degree of relatedness

between two DNA segments. It is determined by measuring the ability of the sense strand of one DNA segment to hybridize with the antisense strand of the other DNA segment, under appropriate conditions, to form a double helix. In the double helix, wherever adenine appears in one strand, thymine appears in the other strand. Similarly, wherever guanine is found in one strand, cytosine is found in the other. The greater the relatedness between the nucleotide sequences of two DNA segments, the greater the ability to form hybrid duplexes between the strands of two DNA segments.

The term "identity" refers to the relatedness of two sequences on a nucleotide-by-nucleotide basis over a particular comparison window or segment. Thus, identity is defined as the degree of sameness, correspondence or equivalence between the same strands (either sense or antisense) of two DNA segments (or two amino acid sequences). "Percentage of sequence identity" is calculated by comparing two optimally aligned sequences over a particular region, determining the number of positions at which the identical base or amino acid occurs in both sequences in order to yield the number of matched positions, dividing the number of such positions by the total number of positions in the segment being compared and multiplying the result by 100. Optimal alignment of sequences may be conducted by the algorithm of Smith & Waterman, Appl. Math. 2:482 (1981), by the algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the method of Pearson & Lipman, *Proc. Natl. Acad. Sci. (USA)* 85:2444 (1988) and by computer programs which implement the relevant algorithms (e.g., Clustal Macaw Pileup (http://cmgm.stanford.edu/biochem218/11Multiple.pdf; Higgins et al., CABIOS. 5L151-153 (1989)), FASTDB

(Intelligenetics), BLAST (National Center for Biomedical Information; Altschul et al., Nucleic Acids Research 25:3389-3402 (1997)), PILEUP (Genetics Computer Group, Madison, WI) or GAP, BESTFIT, FASTA and TFASTA (Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, Madison, WI). (See U.S. Patent No. 5,912,120.)

"Identity between two amino acid sequences is defined as the presence of a series of exactly alike or invariant amino acid residues in both sequences (see above definition for identity between nucleic acid sequences). The definitions of "complementarity" and "identity" are well known to those of ordinary skill in the art.

"Encoded by" refers to a nucleic acid sequence which codes for a polypeptide sequence, wherein the polypeptide sequence or a portion thereof contains an amino acid sequence of at least 3 amino acids, more preferably at least 8 amino acids, and even more preferably at least 15 amino acids from a polypeptide encoded by the nucleic acid sequence.

The present invention also encompasses an isolated nucleic sequence which encodes a protein having polyketide synthase activity and that is hybridizable, under moderately stringent conditions, to a nucleic acid having a nucleotide sequence comprising or complementary to the nucleotide sequences described above. A nucleic acid molecule is "hybridizable" to another nucleic acid molecule when a single-stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and ionic strength (see Sambrook et al., "Molecular Cloning: A Laboratory Manual, Second Edition (1989), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York)). The conditions of temperature and ionic strength

determine the "stringency" of the hybridization. "Hybridization" requires that two nucleic acids contain complementary sequences. However, depending on the stringency of the hybridization, mismatches between bases may occur. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation. Such variables are well known in the art. More specifically, the greater the degree of similarity, identity or homology between two nucleotide sequences, the greater the value of Tm for hybrids of nucleic acids having those sequences. For hybrids of greater than 100 nucleotides in length, equations for calculating Tm have been derived (see Sambrook et al., *supra* (1989)). For hybridization with shorter nucleic acids, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (see Sambrook et al., *supra* (1989)).

As used herein, an "isolated nucleic acid fragment or sequence" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA. (A "fragment" of a specified polynucleotide refers to a polynucleotide sequence which comprises a contiguous sequence of approximately at least about 6 nucleotides, preferably at least about 8 nucleotides, more preferably at least about 10 nucleotides, and even more preferably at least about 15 nucleotides, and most preferable at least about 25 nucleotides identical or complementary to a region of the specified nucleotide sequence.) Nucleotides (usually found in their 5'-monophosphate form) are referred to by

their single letter designation as follows: "A" for adenylate or deoxyadenylate (for RNA or DNA, respectively), "C" for cytidylate or deoxycytidylate, "G" for guanylate or deoxyguanylate, "U" for uridylate, "T" for deoxythymidylate, "R" for purines (A or G), "Y" for pyrimidines (C or T), "K" for G or T, "H" for A or C or T, "I" for inosine, and "N" for any nucleotide.

The terms "fragment or subfragment that is functionally equivalent" and "functionally equivalent fragment or subfragment" are used interchangeably herein. These terms refer to a portion or subsequence of an isolated nucleic acid fragment in which the ability to alter gene expression or produce a certain phenotype is retained whether or not the fragment or subfragment encodes an active enzyme. For example, the fragment or subfragment can be used in the design of chimeric constructs to produce the desired phenotype in a transformed plant. Chimeric constructs can be designed for use in co-suppression or antisense by linking a nucleic acid fragment or subfragment thereof, whether or not it encodes an active enzyme, in the appropriate orientation relative to a plant promoter sequence.

The terms "homology", "homologous", "substantially similar" and " corresponding substantially" are used interchangeably herein. They refer to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate gene expression or produce a certain phenotype. These terms also refer to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially alter the functional properties of the resulting nucleic acid fragment relative to the initial, unmodified fragment. It is therefore

understood, as those skilled in the art will appreciate, that the invention encompasses more than the specific exemplary sequences.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence.

"Native gene" refers to a gene as found in nature with its own regulatory sequences. In contrast,"chimeric construct" refers to a combination of nucleic acid fragments that are not normally found together in nature. Accordingly, a chimeric construct may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that normally found in nature. (The term "isolated" means that the sequence is removed from its natural environment.)

A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric constructs. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include, but are not limited to,

promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" (or "regulatory sequence") refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. The promoter sequence, for example, consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Regulatory sequences (e.g., a promotor) can also be located within the transcribed portions of genes, and/or downstream of the transcribed sequences. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most host cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of some variation may have identical promoter activity.

An "intron" is an intervening sequence in a gene that does not encode a portion of the protein sequence.

Thus, such sequences are transcribed into RNA but are then excised and are not translated. The term is also used for the excised RNA sequences. An "exon" is a portion of the gene sequence that is transcribed and is found in the mature messenger RNA derived from the gene, but is not necessarily a part of the sequence that encodes the final gene product.

The "translation leader sequence" refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner, R. and Foster, G. D. (1995) *Molecular Biotechnology 3*:225).

The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., (1989) *Plant Cell* 1:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from post-transcriptional processing of the primary transcript and is referred to as the mature RNA.

"Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a DNA that is complementary to and synthesized from a mRNA template using the enzyme reverse transcriptase. The cDNA can be single-stranded or converted into the double-stranded form using the Klenow fragment of DNA polymerase I. "Sense" RNA refers to RNA transcript that includes the mRNA and can be translated into protein within a cell or *in vitro*. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Patent No. 5,107,065). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes. The terms "complement" and "reverse complement" are used interchangeably herein with respect to mRNA transcripts, and are meant to define the antisense RNA of the message.

The term "endogenous RNA" refers to any RNA which is encoded by any nucleic acid sequence present in the genome of the host prior to transformation with the recombinant construct of the present invention, whether naturally-occurring or non-naturally occurring, i.e., introduced by recombinant means, mutagenesis, etc.

The term "non-naturally occurring" means artificial, not consistent with what is normally found in nature.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is regulated by the other. For example, a promoter is operably linked with a

coding sequence when it is capable of regulating the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in a sense or antisense orientation. In another example, the complementary RNA regions of the invention can be operably linked, either directly or indirectly, 5' to the target mRNA, or 3' to the target mRNA, or within the target mRNA, or a first complementary region is 5' and its complement is 3' to the target mRNA.

The term "expression", as used herein, refers to the production of a functional end-product. Expression of a gene involves transcription of the gene and translation of the mRNA into a precursor or mature protein. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020).

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or pro-peptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and pro-peptides still present. Pre- and pro-peptides may be but are not limited to intracellular localization signals.

"Stable transformation" refers to the transfer of a nucleic acid fragment into a genome of a host organism, resulting in genetically stable inheritance. In contrast, "transient transformation" refers to the transfer of a nucleic acid fragment into the nucleus, or

DNA-containing organelle, of a host organism resulting in gene expression without integration or stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. The preferred method of cell transformation of rice, corn and other monocots is the use of particle-accelerated or "gene gun" transformation technology (Klein et al., (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050), or an *Agrobacterium*-mediated method using an appropriate Ti plasmid containing the transgene (Ishida et al. (1996) Nature Biotech. 14:745-750). The term "transformation" as used herein refers to both stable transformation and transient transformation.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual;* Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Sambrook").

The term "recombinant" refers to an artificial combination of two otherwise separated segments of sequence, e.g., by chemical synthesis or by the manipulation of isolated segments of nucleic acids by genetic engineering techniques.

"PCR" or "Polymerase Chain Reaction" is a technique for the synthesis of large quantities of specific DNA segments, consists of a series of repetitive cycles (Perkin Elmer Cetus Instruments, Norwalk, CT). Typically, the double stranded DNA is heat denatured, the two primers complementary to the 3' boundaries of the target segment are annealed at low temperature and then extended at an intermediate temperature. One set of these three consecutive steps is referred to as a cycle.

Polymerase chain reaction ("PCR") is a powerful technique used to amplify DNA millions of fold, by repeated replication of a template, in a short period of time. (Mullis et al., *Cold Spring Harbor Symp. Quant. Biol.* 51:263-273 (1986); Erlich et al., European Patent Application 50,424; European Patent Application 84,796; European Patent Application 258,017, European Patent Application 237,362; European Patent Application 201,184, U.S. Patent No. 4,683,202; U.S. Patent No. 4,582,788; and Saiki et al. and U.S. Patent No. 4,683,194). The process utilizes sets of specific in vitro synthesized oligonucleotides to prime DNA synthesis. The design of the primers is dependent upon the sequences of DNA that are to be analyzed. The technique is carried out through many cycles (usually 20-50) of melting the template at high temperature, allowing the primers to anneal to complementary sequences within the template and then replicating the template with DNA polymerase.

The products of PCR reactions are analyzed by separation in agarose gels followed by ethidium bromide staining and visualization with UV transillumination. Alternatively, radioactive dNTPs can be added to the PCR in order to incorporate label into the products. In this case the products of PCR are visualized by exposure of the gel to x-ray film. The added advantage of radiolabeling PCR products is that the levels of individual amplification products can be quantitated.

The terms "recombinant construct", "expression construct" and "recombinant expression construct" are used interchangeably herein. These terms refer to a functional unit of genetic material that can be inserted into the genome of a cell using standard methodology well known to one skilled in the art. Such a construct may be itself or may be used in conjunction with a vector. If a

vector is used, then the choice of vector is dependent upon the method that will be used to transform host plants, as is well known to those skilled in the art. For example, a plasmid can be used. The skilled artisan is well aware of the genetic elements that must be present on the vector in order to successfully transform, select and propagate host cells comprising any of the isolated nucleic acid fragments of the invention. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., (1985) *EMBO J.* 4:2411-2418; De Almeida et al., (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

With respect to "polyketides", these entities are secondary metabolites that are synthesized *via* a series of enzymatic reactions and are analogous to enzymes of the fatty acid synthase (FAS) complex (Hopwood et al., (1990) *Annual Rev. Genet.* 24:37-66). In particular, the enzymes involved in polyketide biosynthesis are called "polyketide synthase enzymes". For purposes herein, "a functionally active polyketide synthase enzyme" is defined as an enzyme or protein involved in the production of polyunsaturated fatty acids such as, for example, eicosapentaenoic acid and docosahexaenoic acid *via* a polyketide-like (PKS-like) pathway (such as described for the production of PUFAs by prokaryotes like Shewanella and Vibrio, and the eukaryote Schizochytrium (see U.S. Patent No. 5,683,898, U.S. Patent No. 6,140,486 and U.S. Patent No. 6,566,583)).

Production of the Polyketide Synthase Enzymes

Once the gene encoding the polyketide synthase enzyme has been isolated, it may then be introduced into either a prokaryotic or eukaryotic host cell, through the use of a vector or construct, in order for the host cell to express the protein of interest. The vector, for example, a bacteriophage, cosmid or plasmid, may comprise the nucleic acid sequence encoding the enzyme, as well as any regulatory sequence (e.g., promoter) that is functional in the host cell and is able to elicit expression of the enzyme encoded by the nucleic acid sequence. The regulatory sequence (e.g., promoter) is in operable association with or operably linked to the nucleotide sequence. (A regulatory sequence (e.g., promoter) is said to be "operably linked" with a coding sequence if the regulatory sequence affects transcription or expression of the coding sequence.) Suitable promoters include, for example, those from genes encoding alcohol dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglucoisomerase, phosphoglycerate kinase, acid phosphatase, T7, TPI, lactase, metallothionein, cytomegalovirus immediate early, whey acidic protein, glucoamylase, promoters activated in the presence of galactose, for example, GAL1 and GAL10, as well as any other promoters involved in prokaryotic and eukaryotic expression systems. Additionally, nucleic acid sequences which encode other proteins, oligosaccharides, lipids, etc., may also be included within the vector as well as other non-promoter regulatory sequences such as, for example, a polyadenylation signal (e.g., the poly-A signal of SV-40T-antigen, ovalalbumin or bovine growth hormone). The choice of sequences present in the construct is dependent

upon the desired expression products as well as the nature of the host cell.

As noted above, once the vector has been constructed, it may then be introduced into the host cell of choice by methods known to those of ordinary skill in the art including, for example, transfection, transformation and electroporation (see Molecular Cloning: A Laboratory Manual, 2nd ed., Vol. 1-3, ed. Sambrook et al., Cold Spring Harbor Laboratory Press (1989)). The host cell is then cultured under suitable conditions permitting expression of the PUFA that is then recovered and purified.

It should also be noted that one may design a unique triglyceride or oil if one utilizes one construct or vector comprising the nucleotide sequences of two or more genes. This vector may then be introduced into one host cell. Alternatively, each of the sequences may be introduced into a separate vector. These vectors may then be introduced into two host cells, respectively, or into one host cell.

Examples of suitable prokaryotic host cells include, for example, bacteria such as Escherichia coli, Bacillus subtilis, Actinomycetes such as Streptomyces coelicolor, Streptomyces lividans, as well as cyanobacteria such as Spirulina spp. (i.e., blue-green algae). Examples of suitable eukaryotic host cells include, for example, mammalian cells, plant cells, yeast cells such as Saccharomyces spp., Lipomyces spp., Candida spp. such as Yarrowia (Candida) spp., Kluyveromyces spp., Pichia spp., Trichoderma spp. or Hansenula spp., or fungal cells such as filamentous fungal cells, for example, Aspergillus, Neurospora and Penicillium. Preferably, Saccharomyces cerevisiae (baker's yeast) cells are utilized.

Expression in a host cell can be accomplished in a transient or stable fashion. Transient expression can occur from introduced constructs which contain expression signals functional in the host cell, but which constructs do not replicate and rarely integrate in the host cell, or where the host cell is not proliferating. Transient expression also can be accomplished by inducing the activity of a regulatable promoter operably linked to the gene of interest, although such inducible systems frequently exhibit a low basal level of expression. Stable expression can be achieved by introduction of a construct that can integrate into the host genome or that autonomously replicates in the host cell. Stable expression of the gene of interest can be selected for through the use of a selectable marker located on or transfected with the expression construct, followed by selection for cells expressing the marker. When stable expression results from integration, the site of the construct's integration can occur randomly within the host genome or can be targeted through the use of constructs containing regions of homology with the host genome sufficient to target recombination with the host locus. Where constructs are targeted to an endogenous locus, all or some of the transcriptional and translational regulatory regions can be provided by the endogenous locus.

A transgenic mammal may also be used in order to express the enzyme of interest (i.e., the polyketide synthase enzyme) encoded by one or both of the above-described nucleic acid sequences. More specifically, once the above-described construct is created, it may be inserted into the pronucleus of an embryo. The embryo may then be implanted into a recipient female. Alternatively, a nuclear transfer method could also be

utilized (Schnieke et al., *Science* (1997) 278:2130-2133). Gestation and birth are then permitted to occur (see, e.g., U.S. Patent No. 5,750,176 and U.S. Patent No. 5,700,671). Milk, tissue or other fluid samples from the offspring should then contain altered levels of PUFAs, as compared to the levels normally found in the non-transgenic animal. Subsequent generations may be monitored for production of the altered or enhanced levels of PUFAs and thus incorporation of the gene or genes encoding the polyketide synthase enzyme into their genomes. The mammal utilized as the host may be selected from the group consisting of, for example, a mouse, a rat, a rabbit, a pig, a goat, a sheep, a horse and a cow. However, any mammal may be used provided it has the ability to incorporate DNA encoding the enzyme of interest into its genome.

For expression of a polyketide synthase polypeptide, functional transcriptional and translational initiation and termination regions are operably linked to the DNA encoding the polypeptide. Transcriptional and translational initiation and termination regions are derived from a variety of nonexclusive sources, including the DNA to be expressed, genes known or suspected to be capable of expression in the desired system, expression vectors, chemical synthesis, or from an endogenous locus in a host cell. Expression in a plant tissue and/or plant part presents certain efficiencies, particularly where the tissue or part is one which is harvested early, such as seed, leaves, fruits, flowers, roots, etc. Expression can be targeted to that location with the plant by utilizing specific regulatory sequence such as those of U.S. Patent Nos. 5,463,174, 4,943,674, 5,106,739, 5,175,095, 5,420,034, 5,188,958, and 5,589,379. Alternatively, the expressed protein can be

an enzyme that produces a product that may be incorporated, either directly or upon further modifications, into a fluid fraction from the host plant. Expression of a polyketide synthase gene or genes, or antisense polyketide synthase transcripts, can alter the levels of specific PUFAs, or derivatives thereof, found in plant parts and/or plant tissues. The polypeptide coding region may be expressed either by itself or with other genes, in order to produce tissues and/or plant parts containing higher proportions of desired PUFAs or in which the PUFA composition more closely resembles that of human breast milk (Prieto et al., PCT publication WO 95/24494). The termination region may be derived from the 3′ region of the gene from which the initiation region was obtained or from a different gene. A large number of termination regions are known to and have been found to be satisfactory in a variety of hosts from the same and different genera and species. The termination region usually is selected as a matter of convenience rather than because of any particular property.

As noted above, a plant (e.g., Glycine max (soybean) or Brassica napus (canola)), plant cell, plant tissue, corn, potato, sunflower, safflower or flax may also be utilized as a host or host cell, respectively, for expression of the polyketide synthase enzyme(s) which may, in turn, be utilized in the production of polyunsaturated fatty acids. More specifically, desired PUFAs can be expressed in seed. Methods of isolating seed oils are known in the art. Thus, in addition to providing a source for PUFAs, seed oil components may be manipulated through the expression of the polyketide synthase genes, in order to provide seed oils that can be added to nutritional compositions, pharmaceutical compositions, animal feeds and cosmetics. Once again, a

vector that comprises a DNA sequence encoding the polyketide synthase enzyme operably linked to a promoter, will be introduced into the plant tissue or plant for a time and under conditions sufficient for expression of the polyketide synthase gene. The vector may also comprise one or more genes which encode other enzymes, for example, elongases, Δ4-desaturase, Δ5-desaturase, Δ6-desaturase, Δ8-desaturase, Δ9-desaturase, Δ10-desaturase, Δ12-desaturase, Δ13-desaturase, Δ15-desaturase, Δ17-desaturase and/or Δ19-desaturase. The plant tissue or plant may produce the relevant substrate (e.g., DGLA, GLA, STA, AA, ADA, EPA, 20:4n-3, etc.) upon which the enzymes act or a vector encoding enzymes which produce such substrates may be introduced into the plant tissue, plant cell, plant, or host cell of interest. In addition, substrate may be sprayed on plant tissues expressing the appropriate enzymes. Using these various techniques, one may produce PUFAs (e.g., n-3 fatty acids such as EPA or DHA) by use of a plant cell, plant tissue, plant, or host cell of interest. It should also be noted that the invention also encompasses a transgenic plant comprising the above-described vector, wherein expression of the nucleotide sequence of the vector results in production of a polyunsaturated fatty acid in, for example, the seeds of the transgenic plant.

The substrates which may be produced by the host cell either naturally, transgenically or exogenously supplied (e.g., acetyl-CoA, malonyl-CoA, malonyl-ACP, methylmalonyl-CoA and methylmalonyl-ACP), as well as the enzymes which may be encoded by DNA sequences introduced in the vector (e.g., polyketide synthase (i.e., β-ketoacyl synthase (or ketoacyl synthase), ketoreductase, dehydratase, and enoyl reductase), which is subsequently introduced into the host cell, in which EPA and/or DHA

is produced. It should be noted that the host cell may produce some of the enzymes (i.e., ketosynthase, ketoreductase, dehydratase and enoyl reductase) endogenously if the PKS genes are expressed individually on different expression vectors.

With respect to the encoded polyketide synthase proteins, it should be noted that the present invention not only encompasses the amino acid sequence of the protein shown in SEQ ID NO:10 but also encompasses proteins comprising amino acid sequences which are at least about 65% identical to, preferably at least about 75% identical to, more preferably at least about 85% identical to and most preferably at least 95% identical to the amino acid sequence shown in SEQ ID NO:10. (All integers within the range of 65 to 100 (in terms of percent identity) are also included within the scope of the invention.) Further, the present invention also encompasses the amino acid sequence of the protein shown in SEQ ID NO:11 as well as all proteins comprising amino acid sequences which are at least about 60% identical to, preferably at least about 70% identical to, more preferably at least about 80% identical to and most preferably at least 90% identical to the amino acid sequence shown in SEQ ID NO:11. (All integers within the range of 60 to 100 (in terms of percent identity) are also included within the scope of the invention.)

In view of the above, the present invention also encompasses a method of producing one or more of the polyketide synthase enzymes described above comprising the steps of: 1) isolating the desired nucleic acid sequence(s) of the gene encoding the synthase(s) (i.e., SEQ ID NO:8 and/or SEQ ID NO:9; 2) constructing a vector comprising said nucleic acid sequence(s); and 3) introducing said vector into a host cell under time and

conditions sufficient for the production of the polyketide synthase enzyme(s).

The present invention also encompasses a method of producing polyunsaturated fatty acids comprising exposing the initial substrates(e.g., acetyl CoA, malonyl CoA, malonyl-ACP, methylmalonyl-CoA and methylmalonyl-ACP) to one or more of the polyketide synthase enzymes described above such that the polyketide synthase converts the initial substrates to a polyunsaturated fatty acid (i.e., EPA or DHA), when additional enzymes are utilized. For example, endogenous acetyl CoA and malonyl CoA (which are found in every cell) are initially condensed by one or more of the polyketide synthases of the present invention. A four-carbon unit fatty acid chain is then formed. In the process, one carbon is lost as carbon dioxide. Subsequently, the four-carbon unit goes through a reduction catalyzed by ketoreductase, dehydration catalyzed by dehydratase, and perhaps another reduction catalyzed by enoyl reductase. Then, the four carbon fatty acid chain is thought to go through repeat cycles and gets extended by two carbons with each cycle until the chain eventually reaches 20 carbon (EPA) or 22 carbons (DHA).

The exact mechanism for the insert of cis double bonds into EPA/DHA is not known but this has been proposed through the action of a bifunctional dehydratase/*2-trans,3-cis* isomerase (DH/2,3I) as seen in E. coli (Metz et al., Science (2001) 293:290-293). Since the PKS cycle extends the chain in two-carbon increments, while the double bond in EPA occurs every third carbon, it has been proposed that the double bonds at carbon atom 14 and carbon atom 8 of EPA are generated by a bifunctional dehydratase/*2-trans,2 cis* isomerase (DH/2,2I). This is followed by the incorporation of a

cis double bond into the elongating fatty acyl chain (Metz et al., Science (2001) 293:290-293).

Uses of the PUFA-Polyketide Synthase Genes and Enzymes Encoded Thereby

As noted above, the isolated nucleic acid sequences (or genes) and the corresponding encoded polyketide synthase enzymes (or purified polypeptides) encoded thereby have many uses. For example, each nucleic acid sequence and corresponding encoded enzyme may be used in the production of polyunsaturated fatty acids, for example, EPA and DHA, as mentioned above. These polyunsaturated fatty acids (i.e., those produced by activity of the polyketide synthase enzyme(s)) may be added to, for example, nutritional compositions, pharmaceutical compositions, cosmetics, and animal feeds, all of which are encompassed by the present invention. Additionally, this system may be used in combination with other genes involved in PUFA biosynthesis such as, for example, the desaturases and elongases involved in DHA production (e.g., Δ4-desaturase and C20-elongase) or related enzymes. Several of these uses are described, in detail, below.

Nutritional Compositions

The present invention includes nutritional compositions. Such compositions, for purposes of the present invention, include any food or preparation for human consumption including for enteral or parenteral consumption, which when taken into the body (a) serve to nourish or build up tissues or supply energy and/or (b) maintain, restore or support adequate nutritional status or metabolic function.

The nutritional composition of the present invention comprises at least one oil or acid produced by use of at least one polyketide synthase enzyme, produced using the respective polyketide synthase gene, and may either be in a solid or liquid form. Additionally, the composition may include edible macronutrients, vitamins and minerals in amounts desired for a particular use. The amount of such ingredients will vary depending on whether the composition is intended for use with normal, healthy infants, children or adults having specialized needs such as those which accompany certain metabolic conditions (e.g., metabolic disorders).

Examples of macronutrients which may be added to the composition include but are not limited to edible fats, carbohydrates and proteins. Examples of such edible fats include but are not limited to coconut oil, soy oil, and mono- and diglycerides. Examples of such carbohydrates include but are not limited to glucose, edible lactose and hydrolyzed starch. Additionally, examples of proteins which may be utilized in the nutritional composition of the invention include but are not limited to soy proteins, electrodialysed whey, electrodialysed skim milk, milk whey, or the hydrolysates of these proteins.

With respect to vitamins and minerals, the following may be added to the nutritional compositions of the present invention: calcium, phosphorus, potassium, sodium, chloride, magnesium, manganese, iron, copper, zinc, selenium, iodine, and Vitamins A, E, D, C, and the B complex. Other such vitamins and minerals may also be added.

The components utilized in the nutritional compositions of the present invention will be of semi-purified or purified origin. By semi-purified or

purified is meant a material which has been prepared by purification of a natural material or by synthesis.

Examples of nutritional compositions of the present invention include but are not limited to infant formulas, dietary supplements, dietary substitutes, and rehydration compositions. Nutritional compositions of particular interest include but are not limited to those utilized for enteral and parenteral supplementation for infants, specialist infant formulae, supplements for the elderly, and supplements for those with gastrointestinal difficulties and/or malabsorption.

The nutritional composition of the present invention may also be added to food even when supplementation of the diet is not required. For example, the composition may be added to food of any type including but not limited to margarines, modified butters, cheeses, milk, yogurt, chocolate, candy, snacks, salad oils, cooking oils, cooking fats, meats, fish and beverages.

In a preferred embodiment of the present invention, the nutritional composition is an enteral nutritional product, more preferably, an adult or pediatric enteral nutritional product. This composition may be administered to adults or children experiencing stress or having specialized needs due to chronic or acute disease states. The composition may comprise, in addition to polyunsaturated fatty acids produced in accordance with the present invention, macronutrients, vitamins and minerals as described above. The macronutrients may be present in amounts equivalent to those present in human milk or on an energy basis, i.e., on a per calorie basis.

Methods for formulating liquid or solid enteral and parenteral nutritional formulas are well known in the art. (See also the Examples below.)

The enteral formula, for example, may be sterilized and subsequently utilized on a ready-to-feed (RTF) basis or stored in a concentrated liquid or powder. The powder can be prepared by spray drying the formula prepared as indicated above, and reconstituting it by rehydrating the concentrate. Adult and pediatric nutritional formulas are well known in the art and are commercially available (e.g., Similac®, Ensure®, Jevity® and Alimentum® from Ross Products Division, Abbott Laboratories, Columbus, Ohio). An oil or fatty acid produced in accordance with the present invention may be added to any of these formulas.

The energy density of the nutritional compositions of the present invention, when in liquid form, may range from about 0.6 Kcal to about 3 Kcal per ml. When in solid or powdered form, the nutritional supplements may contain from about 1.2 to more than 9 Kcals per gram, preferably about 3 to 7 Kcals per gm. In general, the osmolality of a liquid product should be less than 700 mOsm and, more preferably, less than 660 mOsm.

The nutritional formula may include macronutrients, vitamins, and minerals, as noted above, in addition to the PUFAs produced in accordance with the present invention. The presence of these additional components helps the individual ingest the minimum daily requirements of these elements. In addition to the provision of PUFAs, it may also be desirable to add zinc, copper, folic acid and antioxidants to the composition. It is believed that these substance boost a stressed immune system and will therefore provide further benefits to the individual receiving the composition. A pharmaceutical composition may also be supplemented with these elements.

In a more preferred embodiment, the nutritional composition comprises, in addition to antioxidants and at least one PUFA, a source of carbohydrate wherein at least 5 weight % of the carbohydrate is indigestible oligosaccharide. In a more preferred embodiment, the nutritional composition additionally comprises protein, taurine, and carnitine.

As noted above, the PUFAs produced in accordance with the present invention, or derivatives thereof, may be added to a dietary substitute or supplement, particularly an infant formula, for patients undergoing intravenous feeding or for preventing or treating malnutrition or other conditions or disease states. As background, it should be noted that human breast milk has a fatty acid profile comprising from about 0.15% to about 0.36% as DHA, from about 0.03% to about 0.13% as EPA, from about 0.30% to about 0.88% as AA, from about 0.22% to about 0.67% as DGLA, and from about 0.27% to about 1.04% as GLA. Thus, fatty acids such as DGLA, AA, EPA and/or docosahexaenoic acid (DHA), produced in accordance with the present invention, can be used to alter, for example, the composition of infant formulas in order to better replicate the PUFA content of human breast milk or to alter the presence of PUFAs normally found in a non-human mammal's milk. In particular, a composition for use in a pharmacologic or food supplement, particularly a breast milk substitute or supplement, will preferably comprise one or more of AA, DGLA and GLA. More preferably, the oil blend will comprise from about 0.3 to 30% AA, from about 0.2 to 30% DGLA, and/or from about 0.2 to about 30% GLA.

Parenteral nutritional compositions comprising from about 2 to about 30 weight percent fatty acids calculated as triglycerides are encompassed by the present

invention. The preferred composition has about 1 to about 25 weight percent of the total PUFA composition as GLA (U.S. Patent No. 5,196,198). Other vitamins, particularly fat-soluble vitamins such as vitamin A, D, E and L-carnitine can optionally be included. When desired, a preservative such as alpha-tocopherol may be added in an amount of about 0.1% by weight.

In addition, the ratios of AA, DGLA and GLA can be adapted for a particular given end use. When formulated as a breast milk supplement or substitute, a composition which comprises one or more of AA, DGLA and GLA will be provided in a ratio of about 1:19:30 to about 6:1:0.2, respectively. For example, the breast milk of animals can vary in ratios of AA:DGLA:GLA ranging from 1:19:30 to 6:1:0.2, which includes intermediate ratios which are preferably about 1:1:1, 1:2:1, 1:1:4. When produced together in a host cell, adjusting the rate and percent of conversion of a precursor substrate such as GLA and DGLA to AA can be used to precisely control the PUFA ratios. For example, a 5% to 10% conversion rate of DGLA to AA can be used to produce an AA to DGLA ratio of about 1:19, whereas a conversion rate of about 75% to 80% can be used to produce an AA to DGLA ratio of about 6:1. Therefore, whether in a cell culture system or in a host animal, regulating the timing, extent and specificity of elongase expression, as well as the expression of other desaturases, can be used to modulate PUFA levels and ratios. The PUFAs/acids produced in accordance with the present invention (e.g., AA and DGLA) may then be combined with other PUFAs/acids (e.g., GLA) in the desired concentrations and ratios.

Additionally, PUFA produced in accordance with the present invention or host cells containing them may also be used as animal food supplements to alter an animal's

tissue or milk fatty acid composition to one more desirable for human or animal consumption.

Pharmaceutical Compositions

The present invention also encompasses a pharmaceutical composition comprising one or more of the fatty acids and/or resulting oils produced using at least one of the polyketide synthase genes in accordance with the methods described herein. More specifically, such a pharmaceutical composition may comprise one or more of the acids and/or oils as well as a standard, well-known, non-toxic pharmaceutically acceptable carrier, adjuvant or vehicle such as, for example, phosphate buffered saline, water, ethanol, polyols, vegetable oils, a wetting agent or an emulsion such as a water/oil emulsion. The composition may be in either a liquid or solid form. For example, the composition may be in the form of a tablet, capsule, ingestible liquid or powder, injectible, or topical ointment or cream. Proper fluidity can be maintained, for example, by the maintenance of the required particle size in the case of dispersions and by the use of surfactants. It may also be desirable to include isotonic agents, for example, sugars, sodium chloride and the like. Besides such inert diluents, the composition can also include adjuvants, such as wetting agents, emulsifying and suspending agents, sweetening agents, flavoring agents and perfuming agents.

Suspensions, in addition to the active compounds, may comprise suspending agents such as, for example, ethoxylated isostearyl alcohols, polyoxyethylene sorbitol and sorbitan esters, microcrystalline cellulose, aluminum metahydroxide, bentonite, agar-agar and tragacanth or mixtures of these substances.

Solid dosage forms such as tablets and capsules can be prepared using techniques well known in the art. For example, PUFAs produced in accordance with the present invention can be tableted with conventional tablet bases such as lactose, sucrose, and cornstarch in combination with binders such as acacia, cornstarch or gelatin, disintegrating agents such as potato starch or alginic acid, and a lubricant such as stearic acid or magnesium stearate. Capsules can be prepared by incorporating these excipients into a gelatin capsule along with antioxidants and the relevant PUFA(s). The antioxidant and PUFA components should fit within the guidelines presented above.

For intravenous administration, the PUFAs produced in accordance with the present invention or derivatives thereof may be incorporated into commercial formulations such as Intralipids™. The typical normal adult plasma fatty acid profile comprises 6.64 to 9.46% of AA, 1.45 to 3.11% of DGLA, and 0.02 to 0.08% of GLA. These PUFAs or their metabolic precursors can be administered alone or in combination with other PUFAs in order to achieve a normal fatty acid profile in a patient. Where desired, the individual components of the formulations may be provided individually, in kit form, for single or multiple use. A typical dosage of a particular fatty acid is from 0.1 mg to 20 g (up to 100 g) daily and is preferably from 10 mg to 1, 2, 5 or 10 g daily.

Possible routes of administration of the pharmaceutical compositions of the present invention include, for example, enteral (e.g., oral and rectal) and parenteral. For example, a liquid preparation may be administered, for example, orally or rectally. Additionally, a homogenous mixture can be completely dispersed in water, admixed under sterile conditions with

physiologically acceptable diluents, preservatives, buffers or propellants in order to form a spray or inhalant.

The route of administration will, of course, depend upon the desired effect. For example, if the composition is being utilized to treat rough, dry, or aging skin, to treat injured or burned skin, or to treat skin or hair affected by a disease or condition, it may perhaps be applied topically.

The dosage of the composition to be administered to the patient may be determined by one of ordinary skill in the art and depends upon various factors such as weight of the patient, age of the patient, immune status of the patient, etc.

With respect to form, the composition may be, for example, a solution, a dispersion, a suspension, an emulsion or a sterile powder which is then reconstituted.

The present invention also includes the treatment of various disorders by use of the pharmaceutical and/or nutritional compositions described herein. In particular, the compositions of the present invention may be used to treat restenosis after angioplasty. Furthermore, symptoms of inflammation, rheumatoid arthritis, asthma and psoriasis may also be treated with the compositions of the invention. Evidence also indicates that PUFAs may be involved in calcium metabolism; thus, the compositions of the present invention may, perhaps, be utilized in the treatment or prevention of osteoporosis and of kidney or urinary tract stones.

Additionally, the PUFAs produced using the polyketide synthase enzymes of the present invention may also be used in the treatment of cancer. Malignant cells have been shown to have altered fatty acid compositions.

Addition of fatty acids has been shown to slow their growth, cause cell death and increase their susceptibility to chemotherapeutic agents. Moreover, the compositions of the present invention may also be useful for treating cachexia associated with cancer.

The compositions of the present invention may also be used to treat diabetes (see U.S. Patent No. 4,826,877 and Horrobin et al., Am. J. Clin. Nutr. Vol. 57 (Suppl.) 732S-737S). Altered fatty acid metabolism and composition have been demonstrated in diabetic animals.

Furthermore, the compositions of the present invention comprising PUFAs produced either directly or indirectly through the use of the polyketide synthase enzyme(s), may also be used in the treatment of eczema, in the reduction of blood pressure, and in the improvement of mathematics examination scores. Additionally, the compositions of the present invention may be used in inhibition of platelet aggregation, induction of vasodilation, reduction in cholesterol levels, inhibition of proliferation of vessel wall smooth muscle and fibrous tissue (Brenner et al., *Adv. Exp. Med. Biol.* Vol. 83, p.85-101, 1976), reduction or prevention of gastrointestinal bleeding and other side effects of non-steroidal anti-inflammatory drugs (see U.S. Patent No. 4,666,701), prevention or treatment of endometriosis and premenstrual syndrome (see U.S. Patent No. 4,758,592), and treatment of myalgic encephalomyelitis and chronic fatigue after viral infections (see U.S. Patent No. 5,116,871).

Further uses of the compositions of the present invention, the PUFAs of which are produced by use of the polyketide synthase enzymes of the present invention, include use in the treatment of AIDS, multiple sclerosis,

and inflammatory skin disorders, as well as for maintenance of general health.

Veterinary Applications

It should be noted that the above-described PUFA-containing pharmaceutical and nutritional compositions may be utilized in connection with animals (i.e., domestic or non-domestic), as well as humans, as animals experience many of the same needs and conditions as humans. For example, the oil or acids produced using the polyketide synthase enzymes of the present invention may be utilized in animal feed supplements, animal feed substitutes, animal vitamins or in animal topical ointments.

The present invention may be illustrated by the use of the following non-limiting examples:

EXAMPLE I
Construction of BAC library from
Thraustochytrium aureum (ATCC 34304)

Thraustochytrium aureum (ATCC 34304) is an organism that produces copious amounts of polyunsaturated fatty acids (PUFAs) such as DHA which can amount to ~30%-40% of its total fatty acid, a major portion of which appears in its triacylglyceride fraction. This organism belongs to the *Thraustochytrid* family of marine organisms, which include organisms like *Schizochytrium*, *Ulkenia*, *Aplanochytrium* etc, many of which make DHA. Recent studies with *Schizochytrium* have revealed the presence of polyketide synthase (PKS) gene clusters that are involved in DHA biosynthesis (Metz et al., *Science* (2001) 293:290-293; U.S. Patent No. 6,566,583), similar to the PKS gene clusters seen in the EPA- and DHA- producing prokaryotes like *Shewanella* (Yazawa, K., (1996) *Lipids* 31

Suppl.:S297-300) and *Vibrio* (Morita et al., (1999) *Biotechn. Lett.* 21:641-646). Since *Thraustochytrium aureum* and *Schizochytrium* belong to the same family, it was thought that perhaps a similar set of PKS genes might exist in *Thraustochytrium aureum* that are involved in DHA biosynthesis.

To identify the PKS genes involved in EPA/DHA production in *T. aureum*, genomic libraries were constructed in the BAC vectors, TrueBlue-BAC2 (Genomics One, Inc., Quebec, Canada), or pCC1BAC (Epicenter, Madison, WI) and screened with PKS gene probes. For the construction of BAC libraries, high molecular weight genomic DNA was needed. The isolation of this high molecular weight genomic DNA from *T. aureum* was carried out as follows: Frozen fungal pellets were crushed in liquid nitrogen, mixed with Tris-saturated phenol:TE (1:1), and incubated for 10 min at room temperature (RT). The mixture was centrifuged at 6000 rpm for 10 min at RT, after which the aqueous phase was mixed with an equal volume of chloroform: isoamyl alcohol (24:1), and centrifuged as before. The DNA from the aqueous phase thus obtained was precipitated with 0.6 volumes of isopropanol, spun at 13,000 rpm for 20 min, and the pellet thus obtained washed with 70% ethanol, dissolved in TE (pH 8) and then treated with RNase A. The genomic DNA (gDNA) was purified by extractions with phenol:chloroform:isoamyl alcohol (25:24:1), followed by chloroform:isoamyl alcohol (24:1) extraction. The DNA in the aqueous phase was precipitated with 2.5 volumes of ethanol, spun down and washed with ethanol as mentioned earlier. The quality of the isolated gDNA was analyzed by pulsed field gel electrophoresis (PFGE) (CHEF; Amersham Pharmacia, Piscataway, N.J.). The gDNA thus

isolated was ~150-200 Kb in size and did not show much shearing.

The purified gDNA was partially digested using *ClaI* for a time interval of 5 min to 40 min to give a desired size range of 30-40 kb, and digested DNA was separated on a 1.2 % low melting temperature agarose pulse field gel electrophoresis (PFGE) gel. The appropriate sized fractions were excised from the low melting agarose PFGE, eluted from the excised gel, and precipitated using LiCl/Glycogen. The DNA thus obtained was purified by ethanol precipitation as described previously. The size range of the fractions was confirmed on PFGE.

For construction of the BAC library, the TrueBlue-BAC2 vector (Genomics One, Inc., Quebec, Canada) was linearized with *ClaI*, dephosphorylation with Calf Intestinal Alkaline Phosphatase, and ligated to the *ClaI* digested gDNA insert in a molar ratio of 1:5. Ligation was carried out for 16 h at 16°C, followed by transformation into Electromax DH10B *E.coli* competent cells (Invitrogen, Carlsbad, CA). Colonies were grown on selective media containing 25 µg/ml chlormaphenicol, 0.03 mM IPTG and 0.003% Xgal and incubated overnight at 37°C. The average insert size of the library was ~32 kb, library size was 4.8×10^3, with a vector background of 24%.

A BAC library was also constructed in pCC1BAC vector (Epicenter, Madison, WI). Here, the BAC vector was digested with BamHI, dephosphorylation with Calf Intestinal Alkaline Phosphatase, and ligated to the *BamHI* partially digested gDNA insert in a molar ratio of 1:5. Following ligation, EPI300 *E. coli* electrocompetent cells (Epicenter, Madison, WI) were transformed, and transformants grown on selective media containing 12.5 µg/ml Chloramphenicol, 0.4 mM isopropylthiogalactoside

(IPTG) and 40 µg/ml Xgal and incubated overnight at 37°C. The average insert size of the library was ~50 kb, library size was 10^4, with a vector background of 2%.

Example II
Identification of PKS Gene Probes From *Thraustochytrium aureum* (ATCC 34304) for Colony Hybridization

Some of the PKS probes used for the screening of the BAC libraries were identified by random sequencing of a cDNA library constructed from *T. aureum*. The cDNA library was constructed as follows: *T. aureum* (ATCC 34304) cells were grown in BY+ Media (#790, Difco, Detroit, MI) at room temperature for 4 days, in the presence of light, and with constant agitation (250 rpm) to obtain the maximum biomass. These cells were harvested by centrifugation at 5000 rpm for 10 min. and rinsed in ice-cold RNase-free water. These cells were then lysed in a French Press at 10,000 psi, and the lysed cells were directly collected into TE buffered phenol. Proteins from the cell lysate were removed by repeated phenol: chloroform (1:1 v/v) extraction, followed by a chloroform extraction. The nucleic acids from the aqueous phase were precipitated at -70°C for 30 minutes using 0.3M (final concentration) sodium acetate (pH 5.6) and one volume of isopropanol. The precipitated nucleic acids were collected by centrifugation at 15,000 rpm for 30 minutes at 4°C, vacuum-dried for 5 minutes and then treated with DNaseI (RNase-free) in 1X DNase buffer (20 mM Tris-Cl, pH 8.0; 5mM $MgCl_2$) for 15 minutes at room temperature. The reaction was quenched with 5 mM EDTA (pH 8.0) and the RNA further purified using the Qiagen RNeasy Maxi kit (Qiagen, Valencia, CA) as per the manufacturer's protocol.

Messenger RNA (mRNA) was isolated from total RNA using oligo dT cellulose resin, and the pBluescript II XR library construction kit (Stragene, La Jolla, CA) was used to synthesize double stranded cDNA which was then directionally cloned (5' *EcoRI*/ 3' *XhoI*) into pBluescript II SK(+) vector (Stragene, La Jolla, CA). The *T. aureum* library contained approximately 2.5 x 10^6 clones, each with an average insert size of approximately 700 bp.

Random sequencing of this library was carried out on five thousand primary clones which sequenced from the 5' end using the M13 forward primer (5'-AGC GGA TAA CAA TTT CAC ACA GG-3' [SEQ ID NO:1]). Sequencing was carried out using the ABI BigDye sequencing kit (Applied Biosystems, CA) and the MegaBase Capillary DNA sequencer (Amersham Biosciences, Piscataway, NJ). The predicted protein sequences of the library were compared with the predicted protein sequences present in the public database (Genbank) using the NCBI BLASTX program.

Three contigs (Contig 53 [SEQ ID NO:2], Contig 58 [SEQ ID NO:3], and Contig 1763 [SEQ ID NO:4]) were thus identified from the cDNA library sequencing data, which shared homology with regions from published PUFA-PKS genes from *Shewanella* and *Schizochytrium* (Table 1). Sequence comparison of the predicted protein sequences were carried out using the 'BestFit' program in GCG (GCG Wisconsin Package, Madison, WI).

Table 1

Identification of Regions in *T. aureum* With Homology to *Shewanella* PUFA-PKS Genes and *Schizochytrium* PUFA-PKS Genes

Contig #	Length of clone	Location	% Amino Acid Identity (*Shewanella*)	% Amino acid Identity (*Schizochytrium*)
53	713 bp	Region upstream of Ketoacyl reductase (KR) *Shewanella*- ORF 5 *Schizochytrium*- ORF A	41% in 246 aa overlap	36% in 239 aa overlap
58	1023 bp	Region downstream of Ketoacyl reductase (KR) *Shewanella*-ORF 5 *Schizochytrium*- ORF A	32% in 231 aa overlap	43% in 262 aa overlap
1763	1240 bp	Enoyl Reductase (ER) *Shewanella*- ORF 8 *Schizochytrium*- ORF B	52% in 312 aa overlap	75% in 329 aa overlap

Since Contig 53 and Contig 58 were predicted to lie on one open reading frame (ORF) of the PKS cluster, the region between the two contigs which would include the Ketoacyl reductase gene was amplified by PCR using the following primers:

(forward primer) RO 1447 (5'- CTTGTGCAAGAC CTTGGACCTAGAG-3'[SEQ ID NO:5]) based on the sequence of Contig 53;

(reverse primer) RO 1448 (5'-GAACCTCATCCATGTACTGAAACGC-3') [SEQ ID NO:6] based on the sequence of Contig 58.

PCR amplification was carried out using 2 µl of *T. aureum* genomic DNA as a template in a 50 µl total volume containing: PCR buffer [40 mM Tricine-KOH (pH 9.2), 15 mM

KOAc, 3.5 mM $Mg(OAc)_2$, 3.75 µg/ml BSA (final concentration)], 200 µM each deoxyribonucleotide triphosphate, 10 pmole of each primer and 0.5 µl of "Advantage"-brand cDNA polymerase (Clonetech, Palo Alto, CA). Amplification was carried out as follows: initial denaturation at 94°C for 3 minutes, followed by 35 cycles of the following: 94°C for 1 min, 60°C for 30 sec, 72°C for 1 min. A final extension cycle of 72°C for 7 min was carried out, followed by reaction termination at 4°C. The ~1.56 kb PCR product thus produced was labeled 'TA-PKS-1-consensus' or 'TA-PKS-1-1' (SEQ ID NO:7) and was used as a probe for screening the BAC clones to identify clones containing the PKS ORF A region. The predicted protein encoded by TA-PKS 1-1 displayed 52.8% amino acid identity with the homologous region in the *Schizochytrium* ORF A (Figure 1), and 39.9% amino acid identity with the homologous region in ORF 5 of the Shewanella PKS gene cluster (Figure 2), as estimated by using the BestFit program (GCG, Madison, WI). In addition, attempts were made to PCR amplify regions of the PKS cluster corresponding to the β-ketoacyl synthase, malonyl CoA transferase, and the acyl transferase, using degenerate primers that contained conserved motifs shared by PKS genes from Schizochytrium (Metz et al., Science (2001) 293:290-293), Shewanella (Yazawa, K., Lipids (1996) 31 Suppl.: S297-300), Vibrio (Morita et al., Biotechnol. Lett. (1999) 21:641-646) and Photobacterium (Allen et al., Microbiology (2002) 148:1903-1913). However these attempts were unsuccessful.

To identify BAC clones containing the additional sequences present in the PUFA PKS cluster, the Contig 1763 (SEQ ID NO:4) was used as a probe for colony hybridizationto identify clones containing genes

homologous to, for example, the PUFA-PKS genes in ORF 7 and ORF 8 of *Shewanella*. A list of the various probes used for screening the *T. aureum* BAC library is indicated in Table 2.

Table 2

Probes Used for Screening the *T. aureum* Genomic BAC Library by Colony Hybridization

Probe Name	**Probe Length**	**Location on the *Schizochytrium* gene cluster**
TA-PKS-1-1	1560 bp	ORF A
Contig 1763	602 bp	ORF B

Example III

Identification of PUFA-PKS-Related Sequences From *Thraustochytrium aureum*

For screening of the *T. aureum* BAC library with the various probes described above, the library was plated on selective media as described in Example II, and white colonies were replica plated onto Hybond-N+ nylon membranes (Amersham Pharmacia, Piscataway, NJ). The colonies were then lysed by incubation in 10% SDS for 5 min, denatured in [0.5N NaOH,1.5M NaCl] buffer for 5 min, and neutralized in a solution containing [1.5M NaCl,0.5M Tris.Cl (pH 7.4)] for 5 min. The membranes were then incubated in 2X SSC buffer with 0.1% SDS for 5 min, followed by treatment with 0.4 N NaOH for 20 min. Finally the filters were washed once in 2X SSC buffer for 20 min, followed by a wash in 5X SSC buffer for 5 min., and were finally dried at room temperature.

For hybridization, the membranes were prehybridized at 65°C for 10 h in a buffer solution containing [1% BSA, 1mM etylenediaminetetraacedic acid (EDTA)(pH 8.0), 0.5M

NaHPO4 (pH 7.4), 7% SDS, and 10 ug/ml salmon sperm DNA]. Primary hybridization was carried out in 30 ml of the same buffer solution containing DNA probes that were labeled with ^{32}P by random primer labeling using a kit (Stratagene, La Jolla, CA). Specific activity of the probes were $>10^9$ dpm/μg. Hybridization was carried out at 55°C for 16-18 h, which was followed by two washes; the first wash was in a buffer containing [1X SSC + 0.1% SDS] at 55°C for 30 min; the second wash was carried out in a buffer containing [0.1X SSC + 0.1% SDS] at 65°C for 30 min. Membranes were then used to expose X-ray film at -80°C overnight. Positive colonies that were detected by the first screening were subjected to a second round of screening using the same hybridization and washing conditions described above. Colonies selected from the secondary screen were subjected to a PCR screen using primers specific for the probes used, to confirm the presence of the probe sequence in the BAC clones identified.

The TA-PKS 1-1 probe, that contained sequence that was homologous to the ORF A region of the PKS gene cluster in *Schizochytrium* and ORF 5 of the PKS gene cluster in *Shewanella*, was used for screening the BAC library constructed in True-Blue BAC2 vector. This screening resulted in the identification of nine putative positive clones, all of which contained the TA-PKS1-1 probe sequence which was determined by PCR screening. Partial sequencing of three of these nine clones revealed the presence of gene sequences that were homologous to genes present in ORF A and ORF B of the *Schizochytrium* PUFA-PKS gene cluster, as well as homologous to genes present in the ORF 5, ORF 6 and ORF 7 of the *Shewanella* PUFA-PKS gene cluster. Sequences corresponding to those present in ORF C of the *Schizochytrium* PUFA-PKS cluster

or homologous to genes in ORF 8 of *Shewanella* as well as the Dehydratase (DH) genes in ORF 7 of *Shewanella* were not detected in any of these BAC clones. One of these three BAC clones (BAC #164) was selected for full-length sequencing, to determine the entire sequence of the putative PKS gene cluster and also corresponds to sequences present in ORF 5, ORF 6, ORF 7 and ORF 8 of the *Shewanella* PUFA-PKS domains. The full-length sequence of ~50 kb BAC #164 revealed the presence of genes that were organized in the same sequential order as those present in ORF A and ORF B of the *Schizochytrium* PKS gene clusters. The biologically active domains of the *Thraustochytrium aureum* PKS gene cluster are depicted in Figure 3. Details of the domains contained in each ORF are described below.

Thraustochytrium aureum ORFs present on BAC #164

SEQ ID NO:8 ORF A	38,716 to 47,463	8748 bases	Frame1(forward)
SEQ ID NO:9 ORF B	31,128 to 37,250*	6123 bases	Frame2(reverse)

* reverse sequence extending from position 37,250 to 31,128 is shown in SEQ ID NO:9

Open Reading Frame A (ORF A)

The complete nucleotide sequence of ORF A is 8748 bp including the stop codon (SED ID NO:8), and encodes a protein of 2915 amino acids (SEQ ID NO:10). Within ORF A, eleven domains were identified which include:

a. a β-keto-acyl-ACP synthase (KS) domain
b. a malonyl-CoA:ACP acyltransferase (MAT) domain
c. eight acyl carrier protein (ACP) domains
d. a ketoreductase (KR) domain

The sequences of individual domains provided herein are thought to contain the full-length of the sequence

encoding the functional domain, in addition to some flanking regions within the ORF. These domains were identified based on homology comparison with bacterial PUFA-PKS (Metz et al., (2001) *Science* 293:290-293) systems as well as the *Schizochytrium* PUFA-PKS system (Yazawa, K., (1996) 31 Suppl:S297-300). This was done using 'TfastA' (GCG Wisconsin Package, Madison, WI), which uses a method of Pearson and Lipman (Pearson et al., *Proc. Natl. Acad. Sci. USA* (1988) 85:2444-48) to search for similarities between a query peptide sequence and a group of nucleotide sequences translated in all six reading frames. The sequences obtained from *Thraustochytrium aureum* were searched against the GenBank public domain database. In addition, other programs used for analysis include 'BestFit' (GCG Wisconsin Package) which inserts gaps to obtain the optimal alignment of the best region of similarity between two sequences, and 'Gap' (GCG Wisconsin Package) which uses the algorithm of Needleman and Wunsch (*J. Mol. Biol.* (1970) 48:443-53) to align two sequences so as to maximize the number of matches and minimize the number of gaps. In addition, a program Pfam (Bateman et al., (2002) *Nucleic Acids Res.* 30:276-280) was used for analysis. This program can compare proteins or regions of proteins to existing protein domains or conserved protein regions, thus grouping proteins into families based on predicted function.

The domains within ORF A are represented in Table 3.

Table 3

Protein Domains Present in ORF A of the PUFA-PKS genes from *Thraustochytrium aureum*

ORF A Domains	Position on Nucleotide Sequence [@] SEQ ID NO:8	Position on Protein Sequence [@] SEQ ID NO:10	Conserved Motif/Family
KS	289-1764 (SEQ ID NO:12)	97-588 (SEQ ID NO: 13)	DXAC* (*acyl binding site C_{302})
MAT	1975-3305 (SEQ ID NO:14)	659-1101 (SEQ ID NO:15)	GHS*XG (*acyl binding site S_{787})
ACP	3511-3777 (SEQ ID NO:16)[&] 3880-4137 4243-4500 4576-4833 4936-5193 5269-5526 5629-5886 5989-6243	1172-1259 (SEQ ID NO:17) 1295-1380 1415-1501 1527-1611 1648-1732 1758-1843 1878-1962 1997-2082	LGIDS* (*pantetheine binding site S)
KR	6280-8745 (SEQ ID NO:18)	2094-2916 (SEQ ID NO:19)	short chain dehydrogenase family

[@] The actual start and end positions of the domain may be internal to the sequence listed.

[&] The nucleotide and amino acid sequence of the ACP proteins are highly conserved and hence the domain of only one sequence is represented in the sequence identifier.

Open Reading Frame B (ORF B)

The complete nucleic acid sequence of ORF B is 6123 bp (SED ID NO:9) including the stop codon, and encodes a protein of 2040 amino acids (SEQ ID NO:11). Within ORF B, four domains were identified which include:

a. β-keto-acyl-ACP synthase (KS) domain
b. a chain length factor (CLF) domain
c. an acyl transferase (AT) domain
d. an enoyl-ACP-reductase (ER) domain

The domains in ORF B were determined based on homology with the prokaryotic and eukaryotic PUFA-PKS systems as described for ORF A. The sequences of individual domains provided herein are thought to contain the full-length sequence encoding the functional domain, in addition to some flanking regions within the ORF. The domains within ORF B are represented in Table 4.

Table 4

Protein Domains Present in ORF B of the PUFA-PKS Genes From *Thraustochytrium aureum*

ORF B Domains	**Position on Nucleotide Sequence [@] SEQ ID 9**	**Position on Protein Sequence [@] SEQ ID 11**	**Conserved Motif/Family**
KS	79-1461 (SEQ ID NO: 20)	27-487 (SEQ ID NO: 21)	DXAC* (*acyl-binding site C_{237})
CLF	1480-2814 (SEQ ID NO:22)	494-938 (SEQ ID NO: 23)	KS active site motif without acyl-binding cysteine
AT	2815-4302 (SEQ ID NO:24)	939-1434 (SEQ ID NO:25)	GXS*XG (* acyl-binding site S_{1167})
ER	4441-6123 (SEQ ID NO:26)	1481-2041 (SEQ ID NO:27)	

[@] The actual start and end positions of the domain may be internal to the sequence listed.

The overall amino acid sequence comparison of the two ORFs containing the PUFA-PKS genes from *Thraustochytrium aureum* with that of the published *Schizochytrium* PUFA-PKS genes is displayed in Table 5. This sequence comparison was carried out using the 'Gap' program in the GCG Wisconsin package, except where indicated.

Table 5

Comparison of the PUFA-PKS Gene Clusters from *Thraustochytrium aureum* with that from *Schizochytrium* and *Shewanella*

PKS-ORFs Identified from *T. aureum*	**Length of ORFs from *T. aureum***	**% Amino Acid Sequence Identity with Schizochytrium PKS-ORFs**	**% Amino Acid Sequence Identity With Shewanella PUFA-PKS-ORFs**
ORF A	8748 bp	61.1% identity with ORF A	38.4% identity with ORF 5: *KAS domain-49.2% identity *MAT domain-40% identity *ACP domain-~40% identity *KS domain-45% identity
ORF B	6123 bp	59.4% identity with ORF B	21.9% identity with ORF 6: *AT domain-25.8% identity 26% identity with ORF 7: *KS domain-38.3% identity *CLF domain-36.8% identity 48.4% identity with ORF 8: *ER domain-55.2% identity

* Alignments carried out using the "Bestfit" program of GCG.

The functionality of the *Shewanella* PKS gene cluster in generation of long chain PUFAs such as EPA has been well-established (see U.S. Patent No. 5,683,898; Yazawa,

Lipids (1996) 31 Suppl:S297-300; Metz et al., Science (2001) 293:290-293). In addition, sequences from other organisms such as *Vibrio marinus*, which share sequence homology or identity with the *Shewanella* PUFA-PKS genes, have also been shown to be involved in long chain PUFA production (see U.S. Patent No. 6,140,486; Tanaka et al., Biotechnol. Lett. (1999) 21:939). The high sequence homology or identity between the *Thraustochytrium aureum* PKS genes identified herein and the active domains of the *Shewanella* PUFA-PKS gene cluster (see Table 5) indicates that the isolated sequences identified herein have similar functional utility as that of the *Shewanella* and *Vibrio* PKS genes in the production of EPA and DHA.

Example IV

Production of PUFAs in Transgenic Plants

The two ORFs from *Thraustochytrium aureum* may be cloned into suitable plant expression cassettes to be used for plant transformation. Since ORF A and ORF B are within the vicinity of each other, they may be cloned into a single expression cassette in one plant or into two separate expression cassettes in separate plants. If separate plants are used, a heterozygous seed may be produced by crossing the two transgenic plants. Standard transformation protocols may be used which include *Agrobacterium* transformation, or particle bombardment transformation protocols. Transformants may be identified by growing plants on selective media, and transformation of the full-length constructs may be verified by Southern Blot analysis. Immature seeds may also be tested for protein expression of the enzymes encoded by the two ORFs by immunoblotting. The best expressing plants may then be selected and further propagated for further experimentation. The seeds may

also be analyzed for (EPA/DHA) PUFA production, and the best producers grown out and developed through conventional breeding techniques.